

In AI We Trust: Ethics, Artificial Intelligence, and Reliability

Year: 2020 | Citations: 400 | Authors: M. Ryan

Abstract

One of the main difficulties in assessing artificial intelligence (AI) is the tendency for people to anthropomorphise it. This becomes particularly problematic when we attach human moral activities to AI. For example, the European Commission's High-level Expert Group on AI (HLEG) have adopted the position that we should establish a relationship of trust with AI and should cultivate trustworthy AI (HLEG AI Ethics guidelines for trustworthy AI, 2019, p. 35). Trust is one of the most important and defining activities in human relationships, so proposing that AI should be trusted, is a very serious claim. This paper will show that AI cannot be something that has the capacity to be trusted according to the most prevalent definitions of trust because it does not possess emotive states or can be held responsible for their actions—requirements of the affective and normative accounts of trust. While AI meets all of the requirements of the rational account of trust, it will be shown that this is not actually a type of trust at all, but is instead, a form of reliance. Ultimately, even complex machines such as AI should not be viewed as trustworthy as this undermines the value of interpersonal trust, anthropomorphises AI, and diverts responsibility from those developing and using them.